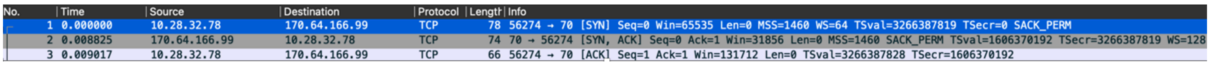
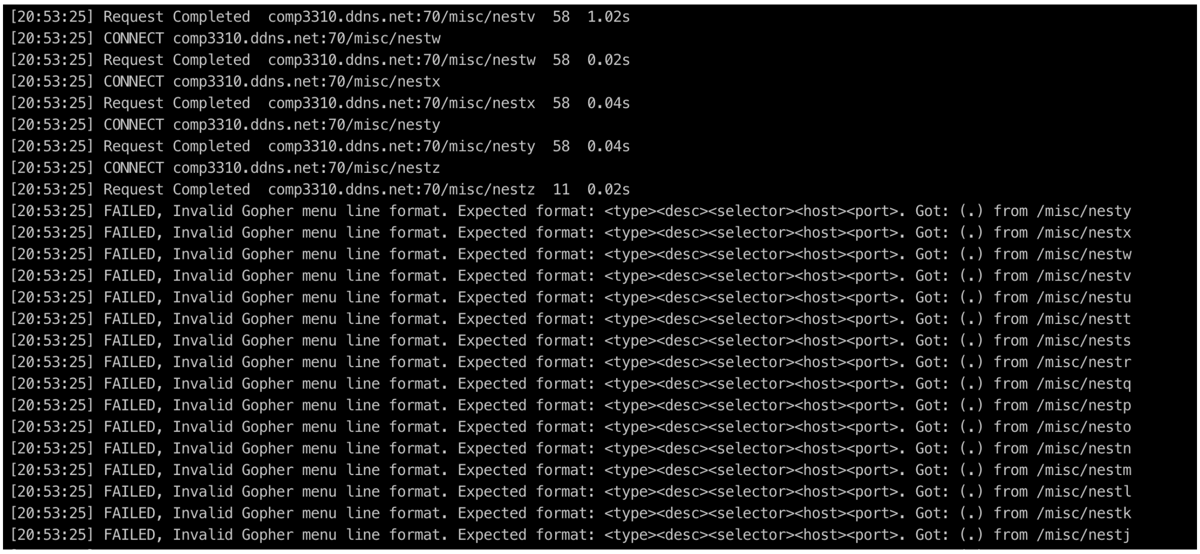
Connect to the server



No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000	10.28.32.78	170.64.166.99	TCP	78	56274 → 70 [SYN] Seq=0 Win=65535 Len=0 MSS=1460 WS=64 TSval=3266387819 TSecr=0 SACK_PERM
2	0.008825	170.64.166.99	10.28.32.78	TCP	74	70 → 56274 [SYN, ACK] Seq=0 Ack=1 Win=31856 Len=0 MSS=1460 SACK_PERM TSval=1606370192 TSecr=3266387819 WS=128
3	0.009017	10.28.32.78	170.64.166.99	TCP	66	56274 → 70 [ACK] Seq=1 Ack=1 Win=131712 Len=0 TSval=3266387828 TSecr=1606370192

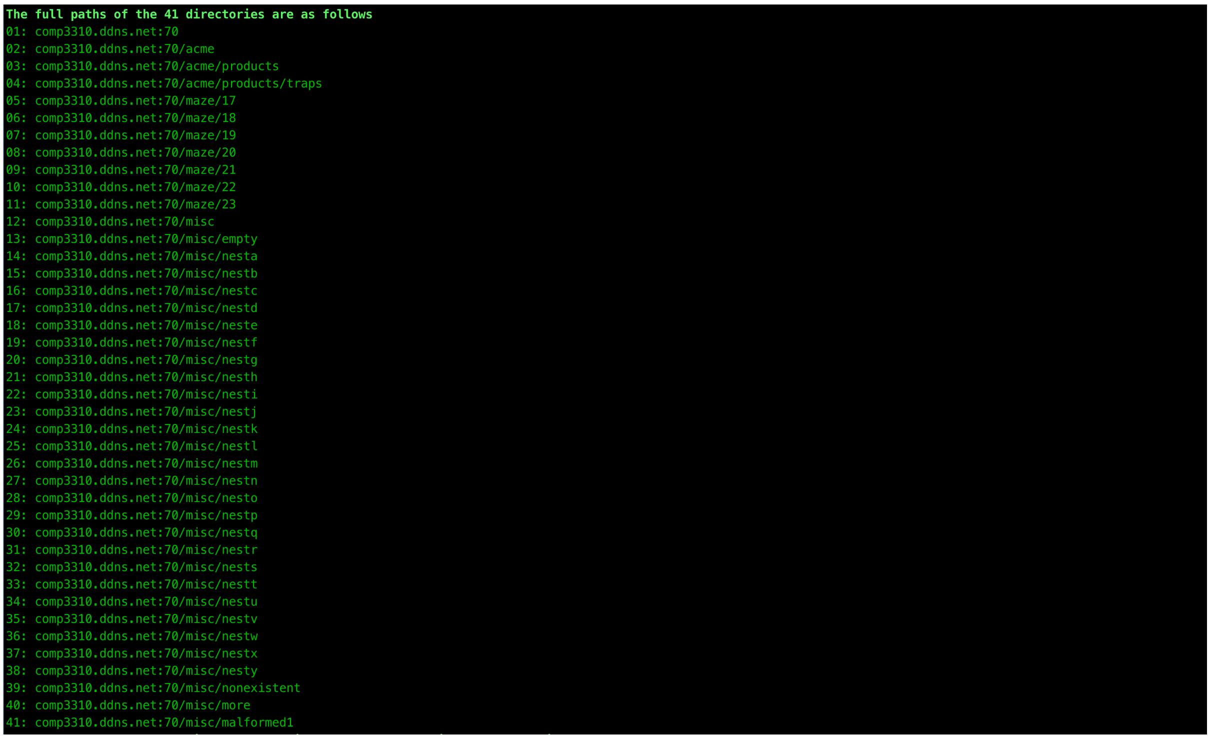
Indexer prints requests and responses with timestamps



```
[20:53:25] Request Completed  comp3310.ddns.net:70/misc/nestv  58  1.02s
[20:53:25] CONNECT comp3310.ddns.net:70/misc/nestw
[20:53:25] Request Completed  comp3310.ddns.net:70/misc/nestw  58  0.02s
[20:53:25] CONNECT comp3310.ddns.net:70/misc/nestx
[20:53:25] Request Completed  comp3310.ddns.net:70/misc/nestx  58  0.04s
[20:53:25] CONNECT comp3310.ddns.net:70/misc/nesty
[20:53:25] Request Completed  comp3310.ddns.net:70/misc/nesty  58  0.04s
[20:53:25] CONNECT comp3310.ddns.net:70/misc/nestz
[20:53:25] Request Completed  comp3310.ddns.net:70/misc/nestz  11  0.02s
[20:53:25] FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nesty
[20:53:25] FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestx
[20:53:25] FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestw
[20:53:25] FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestv
[20:53:25] FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestu
[20:53:25] FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestt
[20:53:25] FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nests
[20:53:25] FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestr
[20:53:25] FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestq
[20:53:25] FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestp
[20:53:25] FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nesto
[20:53:25] FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestn
[20:53:25] FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestm
[20:53:25] FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestl
[20:53:25] FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestk
[20:53:25] FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestj
```

Directories

There are 41 Directories that the indexer was able to find. Their full paths are as follows:



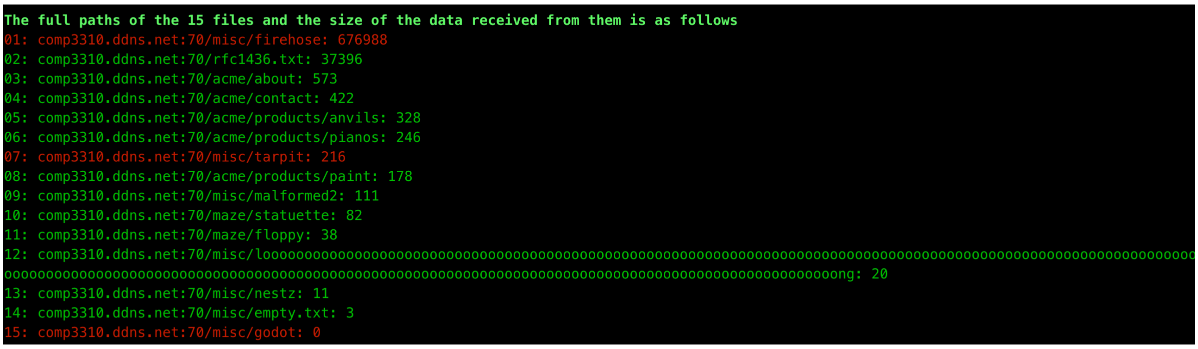
```
The full paths of the 41 directories are as follows
01: comp3310.ddns.net:70
02: comp3310.ddns.net:70/acme
03: comp3310.ddns.net:70/acme/products
04: comp3310.ddns.net:70/acme/products/traps
05: comp3310.ddns.net:70/maze/17
06: comp3310.ddns.net:70/maze/18
07: comp3310.ddns.net:70/maze/19
08: comp3310.ddns.net:70/maze/20
09: comp3310.ddns.net:70/maze/21
10: comp3310.ddns.net:70/maze/22
11: comp3310.ddns.net:70/maze/23
12: comp3310.ddns.net:70/misc
13: comp3310.ddns.net:70/misc/empty
14: comp3310.ddns.net:70/misc/nesta
15: comp3310.ddns.net:70/misc/nestb
16: comp3310.ddns.net:70/misc/nestc
17: comp3310.ddns.net:70/misc/nestd
18: comp3310.ddns.net:70/misc/neste
19: comp3310.ddns.net:70/misc/nestf
20: comp3310.ddns.net:70/misc/nestg
21: comp3310.ddns.net:70/misc/nesth
22: comp3310.ddns.net:70/misc/nesti
23: comp3310.ddns.net:70/misc/nestj
24: comp3310.ddns.net:70/misc/nestk
25: comp3310.ddns.net:70/misc/nestl
26: comp3310.ddns.net:70/misc/nestm
27: comp3310.ddns.net:70/misc/nestn
28: comp3310.ddns.net:70/misc/nesto
29: comp3310.ddns.net:70/misc/nestp
30: comp3310.ddns.net:70/misc/nestq
31: comp3310.ddns.net:70/misc/nestr
32: comp3310.ddns.net:70/misc/nests
33: comp3310.ddns.net:70/misc/nestt
34: comp3310.ddns.net:70/misc/nestu
35: comp3310.ddns.net:70/misc/nestv
36: comp3310.ddns.net:70/misc/nestw
37: comp3310.ddns.net:70/misc/nestx
38: comp3310.ddns.net:70/misc/nesty
39: comp3310.ddns.net:70/misc/nonexistent
40: comp3310.ddns.net:70/misc/more
41: comp3310.ddns.net:70/misc/malformed1
```

The data for each directory is stored using a custom Dir object in Artifacts/Dir.py.

```python
from dataclasses import dataclass
from typing_extensions import Self
from dataclasses import field
from os import makedirs

@dataclass(frozen=True)
class Dir:
  path    : str
  raw     : bytes
  error   : str       = ''
  lines   : list[str] = field(init=False)
  success : bool      = field(init=False)

  def __post_init__(s: Self):
    lines = s.raw.decode("utf-8", "replace").splitlines()
    if lines and lines[-1] == '.': lines.pop()
    object.__setattr__(s, 'lines', s.raw.decode("utf-8", "replace").splitlines())
    object.__setattr__(s, 'success', False if s.error else True)

  def __str__(s: Self)->str: return s.path
```

Text Files

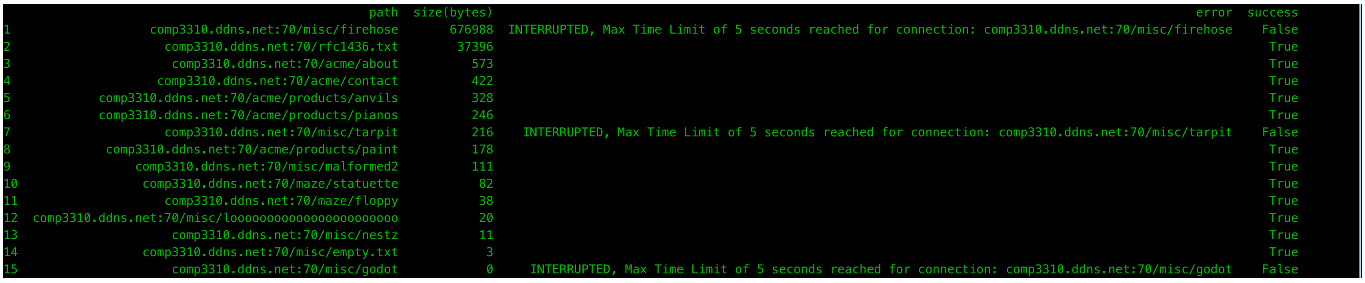
There are 15 text files that the indexer was able to find. Their paths alongside their sizes are as follows. The ones in red either had some problem with the formatting or would cause the client to halt if not handled properly.



```
The full paths of the 15 files and the size of the data received from them is as follows
01: comp3310.ddns.net:70/misc/firehose: 676988
02: comp3310.ddns.net:70/rfc1436.txt: 37396
03: comp3310.ddns.net:70/acme/about: 573
04: comp3310.ddns.net:70/acme/contact: 422
05: comp3310.ddns.net:70/acme/products/anvils: 328
06: comp3310.ddns.net:70/acme/products/pianos: 246
07: comp3310.ddns.net:70/misc/tarpit: 216
08: comp3310.ddns.net:70/acme/products/paint: 178
09: comp3310.ddns.net:70/misc/malformed2: 111
10: comp3310.ddns.net:70/maze/statuette: 82
11: comp3310.ddns.net:70/maze/floppy: 38
12: comp3310.ddns.net:70/misc/looooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooo
oooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooooong: 20
13: comp3310.ddns.net:70/misc/nestz: 11
14: comp3310.ddns.net:70/misc/empty.txt: 3
15: comp3310.ddns.net:70/misc/godot: 0
```

The maximum data received from a text file was from /misc/firehose before the connection was interrupted as the server never stopped sending data for that file.

As you can see, the indexer was able to receive some data for all text files except misc/godot.



```
                                                path  size(bytes)                                                                                    error  success
1                   comp3310.ddns.net:70/misc/firehose       676988  INTERRUPTED, Max Time Limit of 5 seconds reached for connection: comp3310.ddns.net:70/misc/firehose    False
2                     comp3310.ddns.net:70/rfc1436.txt        37396                                                                                                      True
3                      comp3310.ddns.net:70/acme/about          573                                                                                                      True
4                    comp3310.ddns.net:70/acme/contact          422                                                                                                      True
5            comp3310.ddns.net:70/acme/products/anvils          328                                                                                                      True
6            comp3310.ddns.net:70/acme/products/pianos          246                                                                                                      True
7                     comp3310.ddns.net:70/misc/tarpit          216    INTERRUPTED, Max Time Limit of 5 seconds reached for connection: comp3310.ddns.net:70/misc/tarpit    False
8             comp3310.ddns.net:70/acme/products/paint          178                                                                                                      True
9                 comp3310.ddns.net:70/misc/malformed2          111                                                                                                      True
10                 comp3310.ddns.net:70/maze/statuette           82                                                                                                      True
11                    comp3310.ddns.net:70/maze/floppy           38                                                                                                      True
12  comp3310.ddns.net:70/misc/looooooooooooooooooooooo           20                                                                                                      True
13                     comp3310.ddns.net:70/misc/nestz           11                                                                                                      True
14                 comp3310.ddns.net:70/misc/empty.txt            3                                                                                                      True
15                     comp3310.ddns.net:70/misc/godot            0     INTERRUPTED, Max Time Limit of 5 seconds reached for connection: comp3310.ddns.net:70/misc/godot    False
```

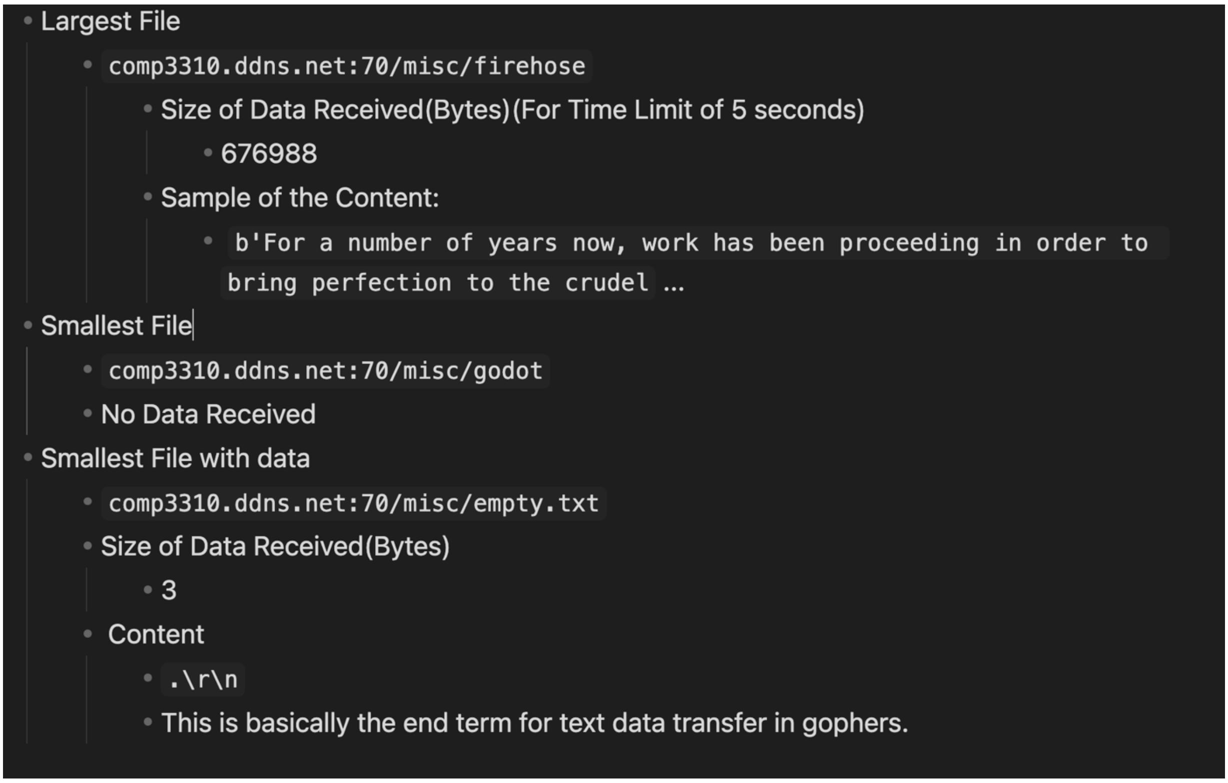
Please also look at the error column in the image above.

If the server terminates connection after successfully sending data, a request is considered successful. The file must also contain '. \r\n' in the end for the indexer to know that it has ended. Otherwise, it will wait until some timeout has occurred. This was done to separate files that end with '. \r\n' from the files that don't.

Requests where no data was received, file did not end with '. \r\n' or the indexer had to interrupt the connection are not considered to be completed successfully. But the indexer stores whatever data was received before some exception is triggered in the TextFile Object.

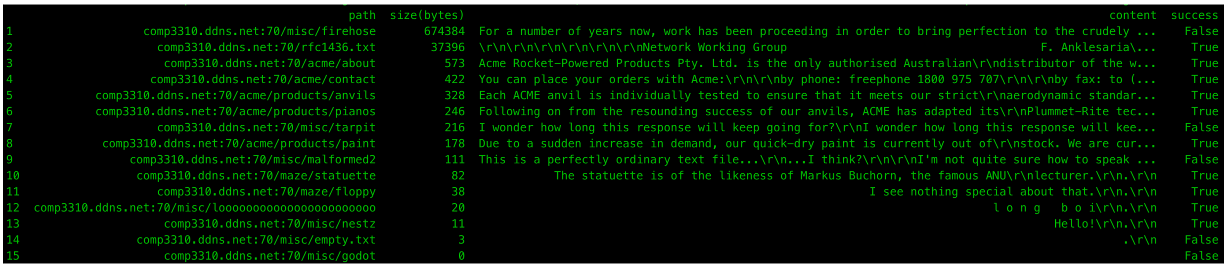
There are different ways that these bad receives from the server are handled, they will be discussed in a later section.

The largest and smallest file findings are as follows



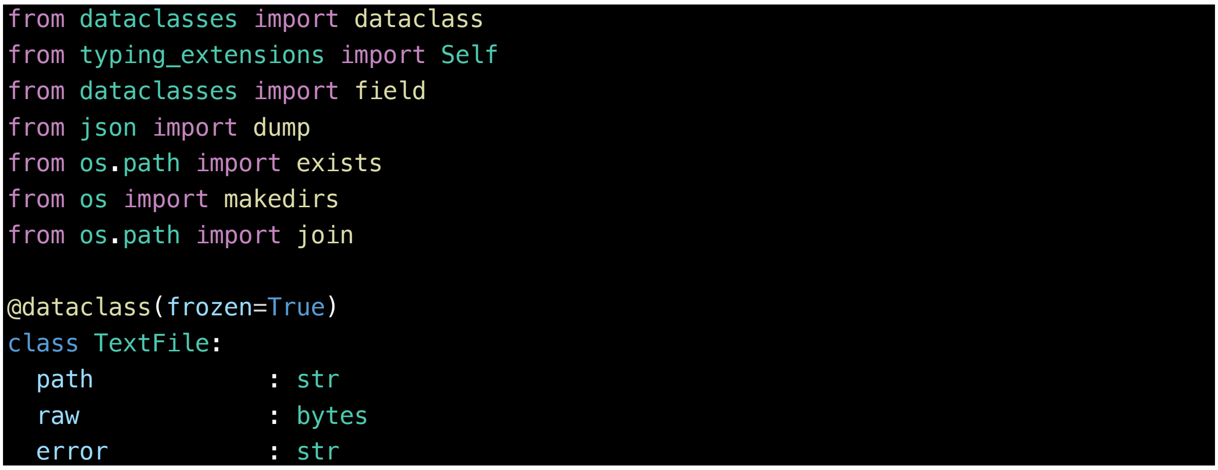
- Largest File
  - `comp3310.ddns.net:70/misc/firehose`
    - Size of Data Received(Bytes)(For Time Limit of 5 seconds)
      - 676988
    - Sample of the Content:
      - `b'For a number of years now, work has been proceeding in order to bring perfection to the crudel` …
- Smallest File
  - `comp3310.ddns.net:70/misc/godot`
  - No Data Received
- Smallest File with data
  - `comp3310.ddns.net:70/misc/empty.txt`
  - Size of Data Received(Bytes)
    - 3
  - Content
    - `.\r\n`
    - This is basically the end term for text data transfer in gophers.

Some more samples of content of each file.

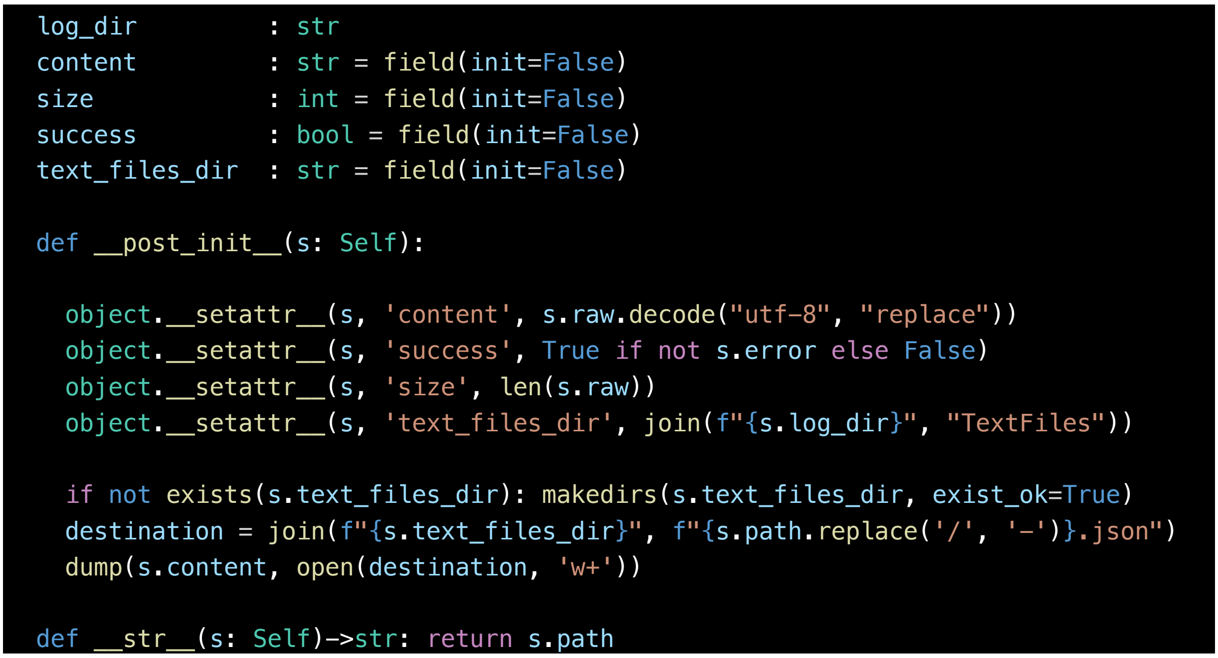


```
                                                path  size(bytes)                                                                                              content  success
1                 comp3310.ddns.net:70/misc/firehose       674384  For a number of years now, work has been proceeding in order to bring perfection to the crudely ...    False
2                   comp3310.ddns.net:70/rfc1436.txt        37396  \r\n\r\n\r\n\r\n\r\n\r\nNetwork Working Group                                     F. Anklesaria\...     True
3                    comp3310.ddns.net:70/acme/about          573  Acme Rocket-Powered Products Pty. Ltd. is the only authorised Australian\r\ndistributor of the w...     True
4                  comp3310.ddns.net:70/acme/contact          422  You can place your orders with Acme:\r\n\r\nby phone: freephone 1800 975 707\r\n\r\nby fax: to (...     True
5          comp3310.ddns.net:70/acme/products/anvils          328  Each ACME anvil is individually tested to ensure that it meets our strict\r\naerodynamic standar...     True
6          comp3310.ddns.net:70/acme/products/pianos          246  Following on from the resounding success of our anvils, ACME has adapted its\r\nPlummet-Rite tec...     True
7                   comp3310.ddns.net:70/misc/tarpit          216  I wonder how long this response will keep going for?\r\nI wonder how long this response will kee...    False
8           comp3310.ddns.net:70/acme/products/paint          178  Due to a sudden increase in demand, our quick-dry paint is currently out of\r\nstock. We are cur...     True
9               comp3310.ddns.net:70/misc/malformed2          111  This is a perfectly ordinary text file...\r\n...I think?\r\n\r\nI'm not quite sure how to speak ...    False
10               comp3310.ddns.net:70/maze/statuette           82             The statuette is of the likeness of Markus Buchorn, the famous ANU\r\nlecturer.\r\n.\r\n     True
11                  comp3310.ddns.net:70/maze/floppy           38                                                           I see nothing special about that.\r\n.\r\n     True
12  comp3310.ddns.net:70/misc/looooooooooooooooooooooo           20                                                                            l o n g   b o i\r\n.\r\n     True
13                   comp3310.ddns.net:70/misc/nestz           11                                                                                     Hello!\r\n.\r\n     True
14               comp3310.ddns.net:70/misc/empty.txt            3                                                                                               .\r\n    False
15                   comp3310.ddns.net:70/misc/godot            0                                                                                                        False
```

The information for each text file is stored in the TextFile object defined in Artifacts/TextFile.py



```python
from dataclasses import dataclass
from typing_extensions import Self
from dataclasses import field
from json import dump
from os.path import exists
from os import makedirs
from os.path import join

@dataclass(frozen=True)
class TextFile:
  path              : str
  raw               : bytes
  error             : str
```



```
log_dir         : str
content         : str = field(init=False)
size            : int = field(init=False)
success         : bool = field(init=False)
text_files_dir  : str = field(init=False)

def __post_init__(s: Self):

  object.__setattr__(s, 'content', s.raw.decode("utf-8", "replace"))
  object.__setattr__(s, 'success', True if not s.error else False)
  object.__setattr__(s, 'size', len(s.raw))
  object.__setattr__(s, 'text_files_dir', join(f"{s.log_dir}", "TextFiles"))

  if not exists(s.text_files_dir): makedirs(s.text_files_dir, exist_ok=True)
  destination = join(f"{s.text_files_dir}", f"{s.path.replace('/', '-')}.json")
  dump(s.content, open(destination, 'w+'))

def __str__(s: Self)->str: return s.path
```

Handling Bad Receives in Text Files

- The indexer puts a cap on the amount of data the server can send. If you change the DEFAULT_TIMEOUT value from 5 to 20 seconds, and keep the MAX_DOWNLOAD_SIZE = 1024 * 1024, you will observe that instead of a timeout error, the indexer will catch a different error related to the maximum download size. So the problem of the sever continuously sending data there is handled using multiple checks.

- For files that do not end with '.\r\n' the indexer collects data but does not exit the receiving loop because the end of file character has not reached. It does leave even if server stops sending information. It only receives after a socket timeout, total receive time timeout, or a thread timeout. This helps distinguish files that follow proper syntax from those that don't.

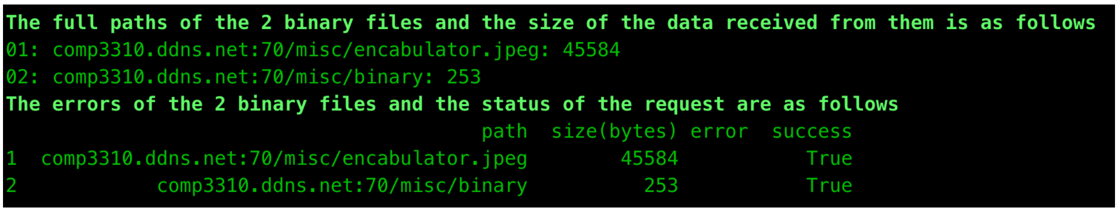
Content

Pleaes check Logs/TextFiles for content of these text files.

Binary Files

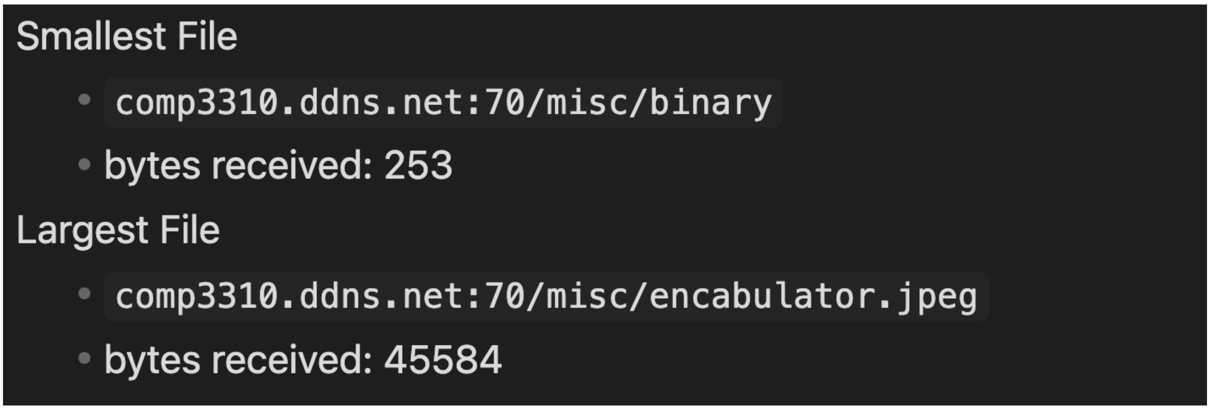
There are 2 Binary Files that the client was able to find. The client managed to fetch the data for both successfully.

Similar to text files, information pertaining to Binary files is stored in the BinaryFiles class in Artifacts/BinaryFile.py



```
The full paths of the 2 binary files and the size of the data received from them is as follows
01: comp3310.ddns.net:70/misc/encabulator.jpeg: 45584
02: comp3310.ddns.net:70/misc/binary: 253
The errors of the 2 binary files and the status of the request are as follows
                                        path  size(bytes) error  success
1  comp3310.ddns.net:70/misc/encabulator.jpeg        45584           True
2            comp3310.ddns.net:70/misc/binary          253           True
```

The findings for smallest and largest binary files are as follows.



Smallest File

- `comp3310.ddns.net:70/misc/binary`
- bytes received: 253

Largest File

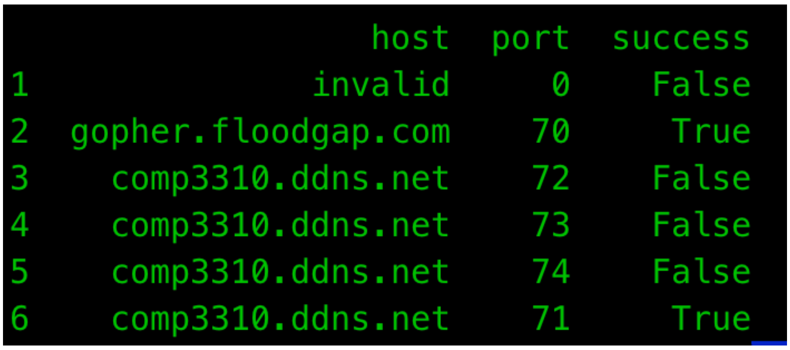
- `comp3310.ddns.net:70/misc/encabulator.jpeg`
- bytes received: 45584

Content

The content can be found in Logs/BinaryFiles

External Servers

6 were found. The findings are as follows

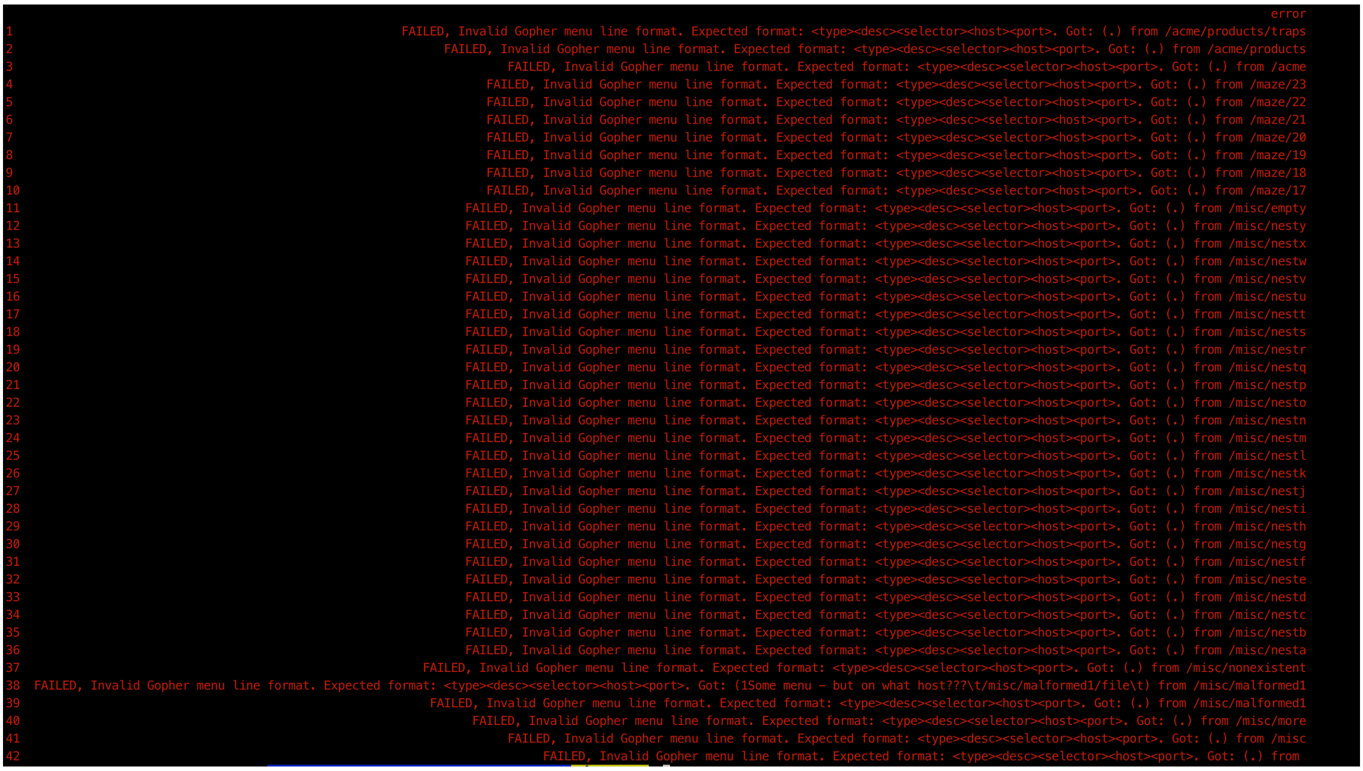


```
                host  port  success
1            invalid     0    False
2  gopher.floodgap.com    70     True
3    comp3310.ddns.net    72    False
4    comp3310.ddns.net    73    False
5    comp3310.ddns.net    74    False
6    comp3310.ddns.net    71     True
```

Bad Directory Lines

Another problem that we encounter is bad directory lines. The client gives the following message after coming across a directory line with improper description

- FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: ({$line}) from {$dir}
- Where ‘$line’ is some line from a directory(‘$dir’).
- There are 42 such cases
- This happens because the line does not correspond to a valid address on the server.



```
                                                                                                                    error
1         FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /acme/products/traps
2               FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /acme/products
3                        FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /acme
4                     FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /maze/23
5                     FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /maze/22
6                     FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /maze/21
7                     FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /maze/20
8                     FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /maze/19
9                     FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /maze/18
10                    FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /maze/17
11                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/empty
12                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nesty
13                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestx
14                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestw
15                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestv
16                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestu
17                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestt
18                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nests
19                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestr
20                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestq
21                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestp
22                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nesto
23                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestn
24                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestm
25                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestl
26                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestk
27                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestj
28                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nesti
29                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nesth
30                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestg
31                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestf
32                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/neste
33                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestd
34                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestc
35                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nestb
36                 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nesta
37           FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/nonexistent
38 FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (1Some menu - but on what host???\t/misc/malformed1/file\t) from /misc/malformed1
39            FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/malformed1
40                  FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc/more
41                       FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from /misc
42                            FAILED, Invalid Gopher menu line format. Expected format: <type><desc><selector><host><port>. Got: (.) from
```

Some more points

1. I have used immutable classes in order to avoid conflicts due to multithreading, although it was probably unnecessary.

2. All the errors that my client was able to come across are explicitly handled.

3. In case the tutor decides to test this on some other gopher, I have added a general exception to the try catch blocks. Hopefully, it works.

4. Tried a different approach with my coding style.

5. I use a mac, hopefully it works on a windows since I am using miniconda.

6. Please check the Artifacts and Summaries folder, I was unable to show them properly in the report.